COFC

R #22

VT


I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450 Alexandria, VA 22313 on October 3, 2003.

Frank C. Eisenschenk, Ph.D., Patent Attorney

REQUEST FOR CERTIFICATE OF CORRECTION UNDER 37 CFR 1.322
Docket No. G-031US03CIP
Patent No. 6,573,068

Certificate
OCT 1 6 2003
of Correction

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | |
|---|---|---|
| Applicants | : | Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Lydie Bougueleret |
| Issued | : | June 3, 2003 |
| Patent No. | : | 6,573,068 |
| For | : | Extended cDNAs for Secreted Proteins |

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR CERTIFICATE OF CORRECTION
UNDER 37 CFR 1.322 (OFFICE MISTAKE)

Sir:

A Certificate of Correction for the above-identified patent has been prepared and is attached hereto.

In the left-hand column below is the column and line number where errors occurred in the patent. In the right-hand column is the page and line number in the application where the correct information appears.

| **Patent Reads:** | **Application Reads:** |
|---|---|
| Column 3, line 62:<br>"as "ESTs" or "5' EST"." | Page 4, lines 11 and 12:<br>--as "ESTs" or "5'ESTs".-- |
| Column 16, line 8:<br>"NOs. 2540 and 4246" | Page 18, line 17:<br>--NOs. 25-40 and 42-46-- |
| Column 18, line 62:<br>"5'-ppgGCAUCCUACUCCCAUCCAAUUCCA CCCAACUCCUCCCAUCUCCAC-3'" | Page 21, Line 29:<br>--5'-pppGCAUCCUACUCCCAUCCAAUUCCA CCCUAACUCCUCCCAUCUCCAC-3'-- |
| Column 22, line 61:<br>"diameter 0.45 mn." | Page 26, line 26:<br>--diameter of 0.45 μm.-- |

OCT 1 6 2003

J:\GEN\031.US03.CIP\PTO-Misc\Req.COC.doc/DNB/tl

| **Patent Reads:** | **Application Reads:** |
|---|---|
| Column 24, line 67:<br>"Caminci" | Page 29, line 9:<br>--Carninci-- |
| Column 28, line 8:<br>"Insert" | Page 32, line 34:<br>--Inserts-- |
| Column 29, line 26:<br>"helix-tum-helix" | Page 34, line 15:<br>--helix-turn-helix-- |
| Column 32, line 19:<br>"14:46834690" | Page 37, line 28:<br>--14:4683-4690-- |
| Column 35, line 55:<br>"270:467470" | Page 41, line 29:<br>--270:467-470-- |
| Column 41, line 43:<br>"46834690" | Page 48, line 24:<br>--4683-4690-- |
| Column 42, line 13:<br>"proteins. 5. Selection of Cloned Full Length Sequence of the Present Invention" | Page 49, lines 12 and 13:<br>--proteins.<br>5. Selection of Cloned Full Length Sequences of the Present Invention-- |
| Column 44, lines 2-3:<br>"Table H" | Page 51, line 18:<br>--Table II-- |
| Column 44, line 7:<br>"Table H" | Page 51, line 21:<br>--Table II-- |
| Column 57, line 50:<br>"11:405411" | Page 67, line 4:<br>--11:405-411-- |
| Column 62, line 7:<br>"13:473486" | Page 72, line 4:<br>--13:473-486-- |
| Column 63, line 59:<br>"tendonaigament formation" | Page 74, line 3:<br>--tendon/ligament formation-- |

| **Patent Reads:** | **Application Reads:** |
|---|---|
| Line 63, line 60:<br>"tendonligament-like" | Page 74, line 4:<br>--tendon/ligament-like-- |
| Column 64, line 11:<br>"tendonaigament cells" | Page 74, line 15:<br>--tendon/ligament cells-- |
| Column 65, line 9:<br>"Chemokincs" | Page 75, line 22:<br>--Chemokines-- |
| Column 70, line 43:<br>"Epitope" | Page 82, line 4:<br>--Epitopes-- |
| Column 74, lines 2-3:<br>"5x10$^{-}$$_{15}$M, and" | Page 86, line 1:<br>--5X10$^{-15}$M, and-- |
| Column 78, line 12:<br>"*Bacic*" | Page 90, line 31:<br>--*Basic*-- |
| Column 80, line 61:<br>"calorimetrically" | Page 94, line 6:<br>--colorimetrically-- |
| Column 85, line 9:<br>"Genomics II1:701-708" | Page 99, line 3:<br>--Genomics 11:701-708-- |
| Column 86, line 63:<br>"Extended cDNA" | Page 101, line 13:<br>--Extended cDNAs-- |
| Column 88, line 30:<br>"Table VI" | Page 103, line 6:<br>--Table VII-- |
| Column 90, line 43:<br>"3 min -94° C." | Page 105, line 5:<br>--3 min -67° C.-- |
| Column 92, line 43:<br>"Protein Which Internet" | Page 107, line 33:<br>--Proteins Which Interact-- |
| Column 98, line 25:<br>"Table VI" | Page 114, line 19:<br>--Table VII-- |
| Column 99, line 18:<br>"412415" | Page 115, line 22:<br>--412-415--. |

| **Patent Reads:** | **Application Reads:** |
|---|---|
| Column 99, line 19:<br>"482488" | Page 115, line 22:<br>--482-488-- |
| Column 99, line 49:<br>"Protein of" | Page 116, line 8:<br>--Proteins of-- |
| Column 100, line 35:<br>"109:429435" | Page 117, line 7:<br>--109:429-435-- |
| Column 103, line 36:<br>"E-cadhenn" | Page 120, line 27:<br>--E-cadherin-- |
| Column 107, line 12:<br>"both 15 in vitro and" | Page 125, lines 14 and 15:<br>--both in vitro and-- |
| Column 108, line 46:<br>"11:428442" | Page 127, line 12:<br>--11:428-442--. |

A true and correct copy of pages 4, 18, 21, 26, 29, 32, 34, 37, 41, 48, 49, 51, 67, 72, 74, 75, 82, 86, 90, 94, 99, 101, 103, 105, 107, 114, 115, 116, 117, 120, 125, and 127 of the specification as filed which supports Applicant's assertion of the error on the part of the Patent Office accompanies this Certificate of Correction.

Approval of the Certificate of Correction is respectfully requested.

Respectfully submitted,

Frank C. Eisenschenk, Ph.D.
Patent Attorney
Registration No. 45,332
Phone No.: 352-375-8100
Fax No.: 352-372-5800
Address: 2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606-6669

FCE/tl

Attachments: Certificate of Correction; copies of pages 4, 18, 21, 26, 29, 32, 34, 37, 41, 48, 49, 51, 67, 72, 74, 75, 82, 86, 90, 94, 99, 101, 103, 105, 107, 114, 115, 116, 117, 120, 125, and 127 of the specification

include sequences adjacent to the sequences of the ESTs may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5′ coding sequences of genes encoding secretory proteins.

## Summary of the Invention

The present invention relates to purified, isolated, or recombinant cDNAs which encode secreted proteins or fragments thereof. Preferably, the purified, isolated or recombinant cDNAs contain the entire open reading frame of their corresponding mRNAs, including a start codon and a stop codon. For example, the cDNAs may include nucleic acids encoding the signal peptide as well as the mature protein. Such cDNAs will be referred herein as "full-length" cDNAs. Alternatively, the cDNAs may contain a fragment of the open reading frame. Such cDNAs will be referred herein as "ESTs" or "5'ESTs". In some embodiments, the fragment may encode only the sequence of the mature protein. Alternatively, the fragment may encode only a fragment of the mature protein. A further aspect of the present invention is a nucleic acid which encodes the signal peptide of a secreted protein.

The present extended cDNAs were obtained using ESTs which include sequences derived from the authentic 5′ ends of their corresponding mRNAs. As used herein the terms "EST" or "5′ EST" refer to the short cDNAs which were used to obtain the extended cDNAs of the present invention. As used herein, the term "extended cDNA" refers to the cDNAs which include sequences adjacent to the 5′ EST used to obtain them. The extended cDNAs may contain all or a portion of the sequence of the EST which was used to obtain them. The term "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5′ EST. As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual extended cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The extended cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately $10^4$-$10^6$ fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "purified" is further used herein to describe a polypeptide or polynucleotide of the invention which has been separated from other compounds including, but not limited to, polypeptides

a fragment thereof of at least 15 consecutive nucleotides. In one aspect of this embodiment, the array includes at least two of the sequences of SEQ ID NOs: 134-180 or 228, the sequences complementary to the sequences of SEQ ID NOs: 134-180 or 228, or fragments thereof of at least 15 consecutive nucleotides. In another aspect of this embodiment, the array includes at least five of the sequences of SEQ ID NOs: 134-180 or 228, the sequences complementary to the sequences of SEQ ID NOs: 134-180 or 228, or fragments thereof of at least 15 consecutive nucleotides.

A further embodiment of the invention encompasses purified polynucleotides comprising an insert from a clone deposited in ATCC accession No. 98619 or a fragment thereof comprising a contiguous span of at least 8, 10, 12, 15, 20, 25, 40, 60, 100, or 200 nucleotides of said insert. An additional embodiment of the invention encompasses purified polypeptides which comprise, consist of, or consist essentially of an amino acid sequence encoded by the insert from a clone deposited in ATCC accession No. 98619, as well as polypeptides which comprise a fragment of said amino acid sequence consisting of a signal peptide, a mature protein, or a contiguous span of at least 5, 8, 10, 12, 15, 20, 25, 40, 60, 100, or 200 amino acids encoded by said insert.

An additional embodiment of the invention encompasses purified polypeptides which comprise, consist of, or consist essentially of an amino acid sequence encoded by the insert from a clone deposited in an ATCC deposit, which contains the sequences of SEQ ID NOs. 25-40 and 42-46, having an accession No. 99061735 and named SignalTag 15061999 or deposited in an ATCC deposit having an accession No. 98121805 and named SignalTag 166-191, which contains SEQ ID NOs.: 47-73, as well as polypeptides which comprise a fragment of said amino acid sequence consisting of a signal peptide, a mature protein, or a contiguous span of at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 amino acids encoded by said insert.

An additional embodiment of the invention encompasses purified polypeptides which comprise a contiguous span of at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 amino acids of SEQ ID NOs: 181-227, wherein said contiguous span comprises at least one of the amino acid positions which was not shown to be identical to a public sequence in the instant application. Also encompassed by the invention are purified polynucleotides encoding said polypeptides.

Another embodiment of the present invention is a computer readable medium having stored thereon a sequence selected from the group consisting of a cDNA code of SEQID NOs. 134-180 or 228 and a polypeptide code of SEQ ID NOs. 181-227 or 229.

Another embodiment of the present invention is a computer system comprising a processor and a data storage device wherein the data storage device has stored thereon a sequence selected from the group consisting of a cDNA code of SEQID NOs. 134-180 or 228 and a polypeptide code of SEQ ID NOs. 181-227 or 229. In some embodiments the computer system further comprises a sequence

## EXAMPLE 1

Ligation of the Nucleoside Diphosphate pCp to the 3′ End of Messenger RNA

1 μg of RNA was incubated in a final reaction medium of 10 μl in the presence of 5 U of $T_4$ phage RNA ligase in the buffer provided by the manufacturer (Gibco - BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2 μl of $^{32}$pCp (Amersham #PB 10208). The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2′, 3′-cis diol at the 3′ ribose, the 2′, 3′-cis diol present at the 5′ end of the mRNA may be oxidized using reagents such as $NaBH_4$, $NaBH_3CN$, or sodium periodate, thereby converting the 2′, 3′-cis diol to a dialdehyde. Example 2 describes the oxidation of the 2′, 3′-cis diol at the 5′ end of the mRNA with sodium periodate.

## EXAMPLE 2

Oxidation of 2′, 3′-cis diol at the 5′ End of the mRNA

0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by in vitro transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the in vitro transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap, m7G(5′)ppp(5′)G. This compound, recognized by polymerase, was incorporated into the 5′ end of the nascent transcript during the step of initiation of transcription but was not capable of incorporation during the extension step. Consequently, the resulting RNA contained a cap at its 5′ end. The sequences of the oligoribonucleotides produced by the in vitro transcription reaction were:

+Cap:

5′m7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3′ (SEQ ID NO:1)

-Cap:

5′-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3′ (SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 μl of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 μl of freshly prepared 0.1 M sodium periodate solution. The mixture was incubated for 1 hour in the dark at 4°C or room temperature. Thereafter, the reaction was stopped by adding 4 μl of 10% ethylene glycol. The product was ethanol precipitated, resuspended in 10μl or more of water or appropriate buffer and dialyzed against water.

prepared in dimethylformamide/water (75:25) containing 2 μg of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C. The mixture was then precipitated twice in $LiClO_4$/acetone. The pellet was resuspended in 200 μl of 0.25 M hydrazine and incubated at 8°C from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in $LiClO_4$/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm which had been stored at -80°C. The mRNA was extracted using conventional acidic phenol techniques. Oligo-dT chromatography was used to purify the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7 μg of the placental mRNAs were oxidized as described above in Example 9. The derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step was replaced by an exclusion chromatography step to remove derivatized oligodeoxyribonucleotides which were not joined to mRNAs. Exclusion chromatography was performed as follows:

10 ml of AcA34 (BioSepra#230151) gel were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

10 μl of the mRNA which had been reacted with the derivatized oligonucleotide were mixed in 39 μl of 10 mM urea and 2 μl of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a filter with a filter of diameter 0.45 μm.

The column was loaded. As soon as the sample had penetrated, equilibration buffer was added. 100 μl fractions were collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Fractions 3 to 15 were combined and precipitated with ethanol.

The mRNAs which had been reacted with the derivatized oligonucleotide were spotted on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The radioactive probe used in these hybridizations was an oligodeoxyribonucleotide of sequence TAATGGTCTCGTGCGAATTCTTGAT (SEQ ID NO:4) which was anticomplementary to the derivatized oligonucleotide and was labeled at its 5′ end with $^{32}P$. 1/10th of the mRNAs which had been

derived. In one version of such procedures, the 5′ ends of the mRNAs are modified as described above. Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5′ end of the mRNA. Single stranded RNAs are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5′ end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule capable of interacting with the molecule used to derivatize the 5′ end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5′ end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci, P. et al. High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper. Genomics 37:327-336 (1996), may be employed to select cDNAs which include the sequence derived from the 5′ end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5′ cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5′ end of the mRNA. Following elimination of the RNA component of the resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a primer complementary to the oligonucleotide tag.

Figure 1 summarizes the above procedures for obtaining cDNAs which have been selected to include the 5′ ends of the mRNAs from which they are derived.

B. Enzymatic Methods for Obtaining mRNAs having Intact 5′ Ends

Other techniques for selecting cDNAs extending to the 5′ end of the mRNA from which they are derived are fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne-Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato et al. Construction of a Human Full-Length cDNA Bank. Gene 150:243-250 (1994).

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate groups present on the 5′ ends of uncapped incomplete mRNAs. Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3′ end, is then ligated to the phosphate present at the 5′ end of the decapped mRNA using T4 RNA ligase. The oligonucleotide may include a restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

## EXAMPLE 12

### Enzymatic Approach for Obtaining 5′ ESTs

Consequently, only the EcoRI site in the oligonucleotide tag was susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra). Fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were selected as described in Example 16 below.

## EXAMPLE 16

### Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang et al., Gene 127:95-8, (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry et al., Biotechniques, 13: 124-131 (1992). In this procedure, the single stranded DNA was hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL) may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

## EXAMPLE 17

### Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer), using

The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data.

Once the sequence data has been stored it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL), BLAST, and BLAST2 program series (NCBI), basic local alignment search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, J. Mol. Biol. 215: 403 (1990)) and FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444 (1988)). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Before searching the cDNAs in the NETGENE™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

## EXAMPLE 18

### Elimination of Undesired Sequences from Further Consideration

5′ ESTs in the NETGENE™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, procaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table I.

To eliminate 5′ ESTs encoding tRNAs from further consideration, the 5′ EST sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5′ ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified as tRNAs and eliminated from further consideration.

To eliminate 5′ ESTs encoding rRNAs from further consideration, the 5′ EST sequences were compared to the sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were

continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

## EXAMPLE 21

### Clustering of the 5′ ESTs and Calculation of Novelty Indices for cDNA Libraries

For each sequenced EST library, the sequences were clustered by the 5′ end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S=107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5′ end were grouped. The longest sequence found in the cluster was used as representative of the cluster. A global clustering between libraries was then performed leading to the definition of super-contigs.

To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as: NR= 100 X (Number of new unique sequences found in the library/Total number of sequences from the library). Typically, novelty rating range between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the libraries, the random sequencing of 5′ EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5′ ESTs in NETGENE™ was screened to identify those 5′ ESTs bearing potential signal sequences as described in Example 22 below.

## EXAMPLE 22

### Identification of Potential Signal Sequences in 5′ ESTs

The 5′ ESTs in the NETGENE™ database were screened to identify those having an uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NETGENE™ contained such an ORF. The ORFs of these 5′ ESTs were searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, G. A New Method for Predicting Signal Sequence Cleavage Sites. Nucleic Acids Res.14:4683-4690 (1986). Those 5′ EST sequences encoding a 15 amino acid long stretch with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5′ ESTs which matched a known human mRNA or EST sequence and had a 5′ end more than 20 nucleotides downstream of the known 5′ end were excluded from further analysis. The remaining cDNAs having signal sequences therein were included in a database called SIGNALTAG™.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an "anchoring enzyme," having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5′ or 3′ most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a "tagging endonuclease" is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short "tag" fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the "tagging endonuclease" to generate short "tag" fragments derived from the cDNAs in the second pool. The "tags" resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce "ditags." In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5′ ESTs or extended cDNAs to determine which 5′ ESTs or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5′ ESTs or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (i.e. extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon), extended cDNAs, 5′ ESTs or fragments of the full length cDNAs, extended cDNAs, or 5′ ESTs of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

For example, quantitative analysis of gene expression may be performed with full length cDNAs, extended cDNAs, 5′ ESTs, or fragments thereof in a complementary DNA microarray as described by Schena et al. Science 270:467-470, 1995; Proc. Natl. Acad. Sci. U.S.A. 93:10614-10619 (1996). Full length cDNAs, extended cDNAs, 5′ ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

stretches using BLASTN were identified as repeat sequences and masked in further identification procedures. In addition, clones showing extensive homology to repeats , i.e., matches of either more than 50 nucleotides if the homology was at least 75% or more than 40 nucleotides if the homology was at least 85% or more than 30 nucleotides if the homology was at least 90%, were flagged.

b) Identification of structural features

Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs are subsequently determined as follows.

A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 20 nt of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. Stretches with 100% homology over 6 nucleotides are identified as polyA tails.

To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp preceding the polyA tail are searched for the canonic polyadenylation AAUAAA signal allowing one mismatch to account for possible sequencing errors and known variation in the canonical sequence of the polyadenylation signal.

c) Identification of functional features

Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length fragments beginning with a translation initiation codon and ending with a stop codon. ORFs encoding at least 20 amino acids are preferred.

Each found ORF is then scanned for the presence of a signal peptide in the first 50 amino-acids or, where appropriate, within shorter regions down to 20 amino acids or less in the ORF, using the matrix method of von Heijne (Nuc. Acids Res. 14: 4683-4690 (1986)), the disclosure of which is incorporated herein by reference and the modification described in Example 22.

d) Homology to either nucleotidic or proteic sequences

Sequences of full length extended cDNAs are then compared to known sequences on a nucleotidic or proteic basis.

Sequences of full length extended cDNAs are compared to the following known nucleic acid sequences: vertebrate sequences (Genbank release # GB), EST sequences (Genbank release # GB), patented sequences (Genseqn release GSEQ) and recently identified sequences (Genbank daily release) available at the time of filing. Full length cDNA sequences are also compared to the sequences of a private database (Genset internal sequences) in order to find sequences that have already been identified by applicants. Sequences of full length extended cDNAs with more than 90% homology over 30 nucleotides using either BLASTN or BLAST2N as indicated in Table II are identified as sequences that

have already been described. Matching vertebrate sequences are subsequently examined using FASTA; full length extended cDNAs with more than 70% homology over 30 nucleotides are identified as sequences that have already been described.

ORFs encoded by full length extended cDNAs as defined in section c) are subsequently compared to known amino acid sequences found in Swissprot release CHP, PIR release PIR# and Genpept release GPEPT public databases using BLASTP with the parameter W=8 and allowing a maximum of 10 matches. Sequences of full length extended cDNAs showing extensive homology to known protein sequences are recognized as already identified proteins.

In addition, the three-frame conceptual translation products of the top strand of full length extended cDNAs are compared to publicly known amino acid sequences of Swissprot using BLASTX with the parameter E=0.001. Sequences of full length extended cDNAs with more than 70% homology over 30 amino acid stretches are detected as already identified proteins.

5. Selection of Cloned Full Length Sequences of the Present Invention

Cloned full length extended cDNA sequences that have already been characterized by the aforementioned computer analysis are then submitted to an automatic procedure in order to preselect full length extended cDNAs containing sequences of interest.

a) Automatic sequence preselection

All complete cloned full length extended cDNAs clipped for vector on both ends are considered. First, a negative selection is operated in order to eliminate unwanted cloned sequences resulting from either contaminants or PCR artifacts as follows. Sequences matching contaminant sequences such as vector RNA, tRNA, mtRNA, rRNA sequences are discarded as well as those encoding ORF sequences exhibiting extensive homology to repeats as defined in section 4 a). Sequences obtained by direct cloning using nested primers on 5' and 3' tags (section 1. case a) but lacking polyA tail are discarded. Only ORFs containing a signal peptide and ending either before the polyA tail (case a) or before the end of the cloned 3'UTR (case b) are kept. Then, ORFs containing unlikely mature proteins such as mature proteins which size is less than 20 amino acids or less than 25% of the immature protein size are eliminated.

In the selection of the OFR, priority was given to the ORF and the frame corresponding to the polypeptides described in SignalTag Patents (United States Patent Application Serial Nos: 08/905,223; 08/905,135; 08/905,051; 08/905,144; 08/905,279; 08/904,468; 08/905,134; and 08/905,133). If the ORF was not found among the OFRs described in the SignalTag Patents, the ORF encoding the signal peptide with the highest score according to Von Heijne method as defined in Example 22 was chosen. If the scores were identical, then the longest ORF was chosen.

Sequences of full length extended cDNA clones are then compared pairwise with BLAST after masking of the repeat sequences. Sequences containing at least 90% homology over 30 nucleotides are

SEQ ID NO:21. This cDNA, falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO:22) having a von Heijne score of 5.9.

The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

The full length cDNA of SEQ ID NO:25 was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

The full length cDNA of SEQ ID NO:27 was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

The above procedures were also used to obtain the extended cDNAs of the present invention. 5′ ESTs expressed in a variety of tissues were obtained as described above. The appended sequence listing provides the tissues from which the extended cDNAs were obtained. It will be appreciated that the extended cDNAs may also be expressed in tissues other than the tissue listed in the sequence listing.

5′ ESTs obtained as described above were used to obtain extended cDNAs having the sequences of SEQ ID NOs: 40-86. Table II provides the sequence identification numbers of the extended cDNAs of the present invention, the locations of the full coding sequences in SEQ ID NOs: 40-86 (i.e. the nucleotides encoding both the signal peptide and the mature protein, listed under the heading FCS location in Table II), the locations of the nucleotides in SEQ ID NOs: 40-86 which encode the signal peptides (listed under the heading SigPep Location in Table II), the locations of the nucleotides in SEQ ID NOs: 40-86 which encode the mature proteins generated by cleavage of the signal peptides (listed under the heading Mature Polypeptide Location in Table II), the locations in SEQ ID NOs: 40-86 of stop codons (listed under the heading Stop Codon Location in Table II), the locations in SEQ ID NOs: 40-86 of polyA signals (listed under the heading Poly A Signal Location in Table II) and the locations of polyA sites (listed under the heading Poly A Site Location in Table II).

The polypeptides encoded by the extended cDNAs were screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at http://expasy.hcuge.ch/sprot/prosite.html. Prosite_convert and prosite_scan programs (http://ulrec3.unil.ch/ftpserveur/prosite_scan) were used to find signatures on the extended cDNAs.

For each pattern obtained with the prosite_convert program from the prosite.dat file, the

Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) Current Protocols in Immunology, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience; Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095 (1980); Weinberger et al., Eur. J. Immun. 11:405-411 (1981); Takai et al., J. Immunol. 137:3494-3500 (1986); and Takai et al., J. Immunol. 140:508-512 (1988).

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

## EXAMPLE 33

### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references: Chapter 3 (In Vitro Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) Current Protocols in Immunology, J.E. Coligan et al. Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-24921 (1981); Herrmann et al., J. Immunol. 128:1968-1974 (1982); Handa et al., J. Immunol. 135:1564-1572 (1985); Takai et al., J. Immunol. 137:3494-3500 (1986); Takai et al., J. Immunol. 140:508-512 (1988); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492 (1981); Herrmann et al J. Immunol. 128:1968-1974 (1982); Handa et al., J. Immunol. 135:1564-1572 (1985); Takai et al., J. Immunol. 137:3494-3500 (1986); Bowman et al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512 (1988); Bertagnolli et al., Cellular Immunology 133:327-341 (1991); and Brown et al., J. Immunol. 153:3079-3092 (1994).

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Maliszewski, J. Immunol. 144:3028-3033 (1990); and Mond, J.J. and Brunswick, M. Assays for B Cell Function: In vitro Antibody Production, Vol 1 pp. 3.8.1-3.8.16 Current Protocols in Immunology. J.E. Coligan et al Eds., John Wiley and Sons, Toronto. (1994).

The proteins encoded by the cDNAs may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity

their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Johansson et al. Cellular Biology 15:141-151 (1995); Keller et al., Molecular and Cellular Biology 13:473-486 (1993); and McClanahan et al., Blood 81:2903-2915 (1993).

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Freshney, M.G. Methylcellulose Colony Forming Assays, Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. (1994); Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911 (1992); McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. (1994); Neben et al., Experimental Hematology 22:353-359 (1994); Ploemacher, R.E. Cobblestone Area Forming Cell Assay, Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 1-21, Wiley-Liss, Inc., New York, NY. (1994); Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 163-179, Wiley-Liss, Inc., New York, NY. (1994); and Sutherland, H.J. Long Term Culture Initiating Cell Assay, Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 139-162, Wiley-Liss, Inc., New York, NY. (1994).

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is beneficial. For example, a protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those

destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e., for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium)

tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokinc damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

## EXAMPLE 36

### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Reproductive Hormones or Cell Movement

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Vale et al., Endocrinology 91:562-572 (1972); Ling et al., Nature 321:779-782 (1986); Vale et al., Nature 321:776-779 (1986); Mason et al., Nature 318:659-663 (1985); Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095 (1986). Chapter 6.12 (Measurement of Alpha and Beta Chemokines) Current Protocols in Immunology, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience ; Taub et al. J. Clin. Invest. 95:1370-1376 (1995); Lind et al. APMIS 103:140-146 (1995); Muller et al. Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867 (1994); and Johnston et al. J. of Immunol. 153:1762-1768 (1994).

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones or cell movement are beneficial. For example, a protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of folic stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin $\alpha$ family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these

181-227.

EXAMPLE 40

Epitopes and Antibody Fusions

A preferred embodiment of the present invention is directed to eiptope-bearing polypeptides and epitope-bearing polypeptide fragments. These epitopes may be "antigenic epitopes" or both an "antigenic epitope" and an "immunogenic epitope". An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response in vivo when the polypeptide is the immunogen. On the other hand, a region of polypeptide to which an antibody binds is defined as an "antigenic determinant" or "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, e.g., Geysen, et al. (1983) Proc. Natl. Acad. Sci. USA 81:39984002. It is particularly noted that although a particular epitope may not be immunogenic, it is nonetheless useful since antibodies can be made in vitro to any epitope.

An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more often at least 8-10 such amino acids. In preferred embodiment, antigenic epitopes comprise a number of amino acids that is any integer between 3 and 50. Fragments which function as epitopes may be produced by any conventional means. See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985), further described in U.S. Patent No. 4,631,211. Methods for determining the amino acids which make up an immunogenic epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping, e.g., the Pepscan method described by H. Mario Geysen et al. (1984); Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506. Another example is the algorithm of Jameson and Wolf, Comp. Appl. Biosci. 4:181-186 (1988) (said references incorporated by reference in their entireties). The Jameson-Wolf antigenic analysis, for example, may be performed using the computer program PROTEAN, using default parameters (Version 4.0 Windows, DNASTAR, Inc., 1228 South Park Street Madison, WI).

The epitope-bearing fragments of the present invention preferably comprises 6 to 50 amino acids (i.e. any integer between 6 and 50, inclusive) of a polypeptide of the present invention. Also, included in the present invention are antigenic fragments between the integers of 6 and the full length sequence of the sequence listing. All combinations of sequences between the integers of 6 and the full-length sequence of a polypeptide of the present invention are included. The epitope-bearing fragments may be specified by either the number of contiguous amino acid residues (as a sub-genus) or by specific N-terminal and C-terminal positions (as species) as described above for the polypeptide fragments of the present invention. Any number of epitope-bearing fragments of the

$5X10^{-11}M$, $10^{-11}M$, $5X10^{-12}M$, $10^{-12}M$, $5X10^{-13}M$, $10^{-13}M$, $5X10^{-14}M$, $10^{-14}M$, $5X10^{-15}M$, and $10^{-15}M$.

Antibodies of the present invention have uses that include, but are not limited to, methods known in the art to purify, detect, and target the polypeptides of the present invention including both in vitro and in vivo diagnostic and therapeutic methods. For example, the antibodies have use in immunoassays for qualitatively and quantitatively measuring levels of the polypeptides of the present invention in biological samples (See, e.g., Harlow et al., 1988).

The antibodies of the present invention may be used either alone or in combination with other compositions. The antibodies may further be recombinantly fused to a heterologous polypeptide at the N- or C-terminus or chemically conjugated (including covalent and non-covalent conjugations) to polypeptides or other compositions. For example, antibodies of the present invention may be recombinantly fused or conjugated to molecules useful as labels in detection assays and effector molecules such as heterologous polypeptides, drugs, or toxins. See, e.g., WO 92/08495; WO 91/14438; WO 89/12624; US Patent 5,314,995; and EP 0 396 387.

The antibodies of the present invention may be prepared by any suitable method known in the art. For example, a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. The term "monoclonal antibody" is not limited to antibodies produced through hybridoma technology. The term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. The term "monoclonal antibody" refers to an antibody that is derived from a single clone, including eukaryotic, prokaryotic, or phage clone, and not the method by which it is produced. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technology.

Hybridoma techniques include those known in the art (See, e.g., Harlow et al. 1988); Hammerling, et al, 1981). (Said references incorporated by reference in their entireties). Fab and F(ab')2 fragments may be produced, for example, from hybridoma-produced antibodies by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments).

Alternatively, antibodies of the present invention can be produced through the application of recombinant DNA technology or through synthetic chemistry using methods known in the art. For example, the antibodies of the present invention can be prepared using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface

The antibodies of the invention may be labeled by any one of the radioactive, fluorescent or enzymatic labels known in the art.

Consequently, the invention is also directed to a method for detecting specifically the presence of a polypeptide of the present invention according to the invention in a biological sample, said method comprising the following steps:

a) bringing into contact the biological sample with a polyclonal or monoclonal antibody that specifically binds a polypeptide of the present invention; and

b) detecting the antigen-antibody complex formed.

The invention also concerns a diagnostic kit for detecting in vitro the presence of a polypeptide of the present invention in a biological sample, wherein said kit comprises:

a) a polyclonal or monoclonal antibody that specifically binds a polypeptide of the present invention, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom described above, which can be unmodified or modified to enhance immunogenicity. Effective

to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis et al. Basic Methods in Molecular Biology, (1986), Elsevier Press. pp 62-65).

A panel of probes based on the sequences of the extended cDNAs (or genomic DNAs obtainable therefrom), or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis et al., supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom).

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of extended cDNAs (or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of extended cDNA probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

## EXAMPLE 46

### Dot Blot Identification Procedure

Another technique for identifying individuals using the extended cDNA sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the extended cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with $P^{32}$ using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et al. supra). The $^{32}P$ labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood et al., Proc.

(Raeymaekers et al., Genomics 29:170-178, (1995)), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer et al., Genomics 14:574-584 (1992)) and 13 loci on the long arm of chromosome 5 (Warrington et al., Genomics 11:701-708 (1991)).

## EXAMPLE 50

### Mapping of Extended cDNAs to Human Chromosomes using PCR techniques

Extended cDNAs (or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the extended cDNA sequence (or the sequence of a genomic DNA obtainable therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich, H.A., PCR Technology; Principles and Applications for DNA Amplification. (1992). W.H. Freeman and Co., New York.

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 $\mu$Cu of a $^{32}$P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given extended cDNA (or genomic DNA obtainable therefrom). DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the extended cDNAs (or genomic DNAs obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the extended cDNA (or genomic DNA obtainable therefrom) will yield an amplified fragment. The extended cDNAs (or genomic DNAs obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that

bands are obtained as previously described (Cherif et al., supra.). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular extended cDNA (or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given chromosome.

Once the extended cDNAs (or genomic DNAs obtainable therefrom) have been assigned to particular chromosomes using the techniques described in Examples 49-51 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

## EXAMPLE 52

### Use of Extended cDNAs to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Ramaiah Nagaraja et al. Genome Research 7:210-222, (March, 1997). Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the extended cDNA (or genomic DNA obtainable therefrom) whose position is to be determined. Once an insert has been found which includes the extended cDNA (or genomic DNA obtainable therefrom), the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the extended cDNA (or genomic DNA obtainable therefrom) was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 53 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

## EXAMPLE 53

The secretion vectors of the present invention include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from an extended cDNA (or genomic DNA obtainable therefrom), such as one of the signal sequences in SEQ ID NOs: 134-180 as defined in Table VII above, is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the extended cDNA (or genomic DNA obtainable therefrom). Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and hplc. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the

used. The first nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular extended cDNA or 5′ EST for which the promoter is to be cloned and should have a melting temperature, length, and location in the extended cDNA or 5′ EST which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (6 cycles) / 2 sec - 94°C, 3 min - 67°C (25 cycles) / 5 min - 67°C

The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques. Alternatively, tow or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5′ EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5′ EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5′ EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5′ ESTs with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example 56.

## EXAMPLE 56

### Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5′ ESTs are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5′ ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the

program MatInspector release 2.0, August 1996.

Figure 8 describes the transcription factor binding sites present in each of these promoters. The columns labeled matrices provides the name of the MatInspector matrix used. The column labeled position provides the 5′ position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5′ EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5′ EST. The column labeled "score" provides the MatInspector score found for this site. The column labeled "length" provides the length of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

The promoters and other regulatory sequences located upstream of the extended cDNAs or 5′ ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described in Example 26 above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5′ EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences using the procedures of Examples 55-57, proteins which interact with the promoter may be identified as described in Example 58 below.

## EXAMPLE 58

### Identification of Proteins Which Interact with Promoter Sequences, Upstream Regulatory Sequences, or mRNA

Sequences within the promoter region which are likely to bind transcription factors may be

EXAMPLE 63

Reassembling & Resequencing of Clones

Further study of the clones reported in SEQ ID NOs: 40 to 86 revealed a series of abnormalities. As a result, the clones were resequenced twice, reanalyzed and the open reading frames were reassigned. The corrected nucleotide sequences have been disclosed in SEQ ID NOs: 134 to 180 and 228 and the predicted amino acid sequences for the corresponding polypeptides have also been corrected and disclosed in SEQ ID NOs: 181 to 227 and 229. The corrected sequences have been placed in the Sequence Listing in the same order as the original sequences from which they were derived.

After this reanalysis process a few apparent abnormalities persisted. The sequences presented in SEQ ID NOs: 134, 149, 151, and 164 are apparently unlikely to be genuine full length cDNAs. These clones are missing a stop codon and are thus more probably 3' truncated cDNA sequences. Similarly, the sequences presented in SEQ ID NOs: 145, 155, and 166 may also not be genuine full length cDNAs based on homolgy studies with existing protein sequences. Although both of these sequences encode a potential start methionine each could represent of 5' truncated cDNA.

In addition, after the reassignment of open reading frames for the clones, new open reading frames were chosen in some instances. In case of SEQ ID NOs: 135, 149, 155, 160, 166, 171, and 175 the new open reading frames were no longer predicted to contain a signal peptide.

Table VII provides the sequence identification numbers of the extended cDNAs of the present invention, the locations of the full coding sequences in SEQ ID NOs: 134-180 (i.e. the nucleotides encoding both the signal peptide and the mature protein, listed under the heading FCS location in Table VII), the locations of the nucleotides in SEQ ID NOs: 134-180 which encode the signal peptides (listed under the heading SigPep Location in Table VII), the locations of the nucleotides in SEQ ID NOs: 134-180 which encode the mature proteins generated by cleavage of the signal peptides (listed under the heading Mature Polypeptide Location in Table VII), the locations in SEQ ID NOs: 134-180 of stop codons (listed under the heading Stop Codon Location in Table VII), the locations in SEQ ID NOs: 134-180 of polyA signals (listed under the heading PolyA Signal Location in Table VII) and the locations of polyA sites (listed under the heading PolyA Site Location in Table VII).

Table VIII lists the sequence identification numbers of the polypeptides of SEQ ID NOs: 181-227, the locations of the amino acid residues of SEQ ID NOs: 181-227 in the full length polypeptide (second column), the locations of the amino acid residues of SEQ ID NOs: 181-227 in the signal peptides (third column), and the locations of the amino acid residues of SEQ ID NOs: 181-227 in the mature polypeptide created by cleaving the signal peptide from the full length polypeptide (fourth column). In Table VIII, and in the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1

and the first amino acid of the signal peptide is designated with the appropriate negative number, in accordance with the regulations governing sequence listings.

EXAMPLE 64

Functional Anaysis of Predicted Protein Sequences

It should be noted that the numbering of amino acids in the protein sequences discussed in Figures 9 to 16, and Table VI, the first methionine encountered is designated as amino acid number 1. In the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1 and the first amino acid of the signal peptide is designated with the appropriate negative number, in accordance with the regulations governing sequence listings.

Protein of SEQ ID NO:181

The protein of SEQ ID NO: 181 is encoded by the extended cDNA SEQ ID NO: 134. The protein of SEQ ID NO: 181 is human strictosidine synthase. Strictodine synthase is a key enzyme in the production of, and therefore useful in making, the pharmaceutically important monoterpene indole alkaloids. Pathways for the production of monoterpene indole alkaloids can be reconstructed in various cell types, for example, insect cell cultures as described in Kutchan, T.M. et al. (1994) Phyochemistry 35(2):353-360. Strictodine synthase can also be produced E. coli and its activity measuring using methods described in, for example, Roessner, C.A. et al. (1992) Protein Expr. Purif. 3(4):295-300; Kutchan, T.M. (1989) FEBS Lett. 257(1):127-130; Pennings, E.J. et al. (1989) Anal. Biochem. 176(2):412-415; Walton, N.J. (1987) Anal. Biochem. 163(2):482-488. Preferred fragments of SEQ ID NO: 181and the mature polypeptide encoded by the corresponding human cDNA of the deposited clone are those with strictodine synthase activity. Further preferred are fragments with not less then 100 fold less activity, not less than 10 fold activity, and not less than 5 fold activity when compared to mature protein.

Protein of SEQ ID NO: 183

The protein of SEQ ID NO: 183, encoded by the extended cDNA SEQ ID NO: 136, is human inositol hexakisphophate kinase-2. Inositol hexakisphophate kinase-2 phosphorylates inositol hexakisphosphate (InsP(6)) to diphosphoinositol pentakisphosphate/inositol heptakisphosphate (InsP(7)), a high energy regulator of cellular trafficking. Human inositol hexakisphophate kinase-2 also stimulates the uptake of inorganic phosphate and its products act as energy reserves. Therefore, hexakisphosphate kinase-2 is an ATP synthase, and its product, diphosphoinositol pentakisphosphate, acts as a high-energy phosphate donor. The human inositol hexakisphophate kinase-2 gene may be

transfected into eukaryotic cells (preferably mammalian, yeast, and insect cells) and expressed to increase their growth, viability, and for more efficient secretions of polypeptides, including recombinant polypeptides. Preferred fragments of SEQ ID NO: 183and the corresponding mature polypeptide encoded by the human cDNA of the deposited clone are those with inositol hexakisphophate kinase-2 activity. Further preferred are fragments with not less then 100 fold less activity, not less than 10 fold activity, and not less than 5 fold activity when compared to mature protein.

Proteins of SEQ ID NOs: 185 and 215:

The proteins of SEQ ID NOs: 185 and 215 encoded by the extended cDNA SEQ ID NOs: 138 and 168, respectively, are MEK binding partners. These proteins enhance enzymatic activation of mitogen-activated protein (MAP) kinase cascade. The MAP kinase pathway is one of the important enzymatic cascade that is conserved among all eukaryotes from yeast to human. This kind of pathway is involved in vital functions such as the regulation of growth, differentiation and apoptosis. These proteins are believed to act by facilitating the interaction of the two sequentially acting kinases MEK1 and ERK1 (Schaffer et al., Science, 281:1668-1671 (1998)).

Thus, the proteins of SEQ ID NO: 185 and 215 are involved in regulating protein-protein interaction in the signal transduction pathways. These proteins may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, neurodegenerative diseases, cardiovascular disorders, hypertension, renal injury and repair and septic shock. More specifically, over expression and mutant forms of this gene can serve as markers for cancer, such as ovarian cancer, using the nucleic acid as a probe or by using antibodies directed to the protein. Cells transfected with this gene have increased growth rate.

Protein of SEQ ID NO: 186

The protein of SEQ ID NO: 186, encoded by the extended cDNA SEQ ID NO: 139, is a new claudin named Claudin-50.

Cell adhesion is a complex process that is important for maintaining tissue integrity and generating physical and permeability barriers within the body. All tissues are divided into discrete compartments, each of which is composed of a specific cell type that adheres to similar cell types. Such adhesion triggers the formation of intercellular junctions (i.e., readily definable contact sites on the surfaces of adjacent cells that are adhering to one another), also known as tight junctions, gap junctions, spot desmosomes and belt desmosomes. The formation of such junctions gives rise to physical and permeability barriers that restrict the free passage of cells and other biological substances from one tissue compartment to another. For example, the blood vessels of all tissues are composed of endothelial cells. In order for components in the blood to enter a given tissue

compartment, they must first pass from the lumen of a blood vessel through the barrier formed by the endothelial cells of that vessel. Similarly, in order for substances to enter the body via the gut, the substances must first pass through a barrier formed by the epithelial cells of that tissue. To enter the blood via the skin, both epithelial and endothelial cell layers must be crossed.

The transmembrane component of tight junctions that has been the most studied is occluding. Occludin is believed to be directly involved in cell adhesion and the formation of tight junctions (Furuse et al., J. Cell Sci. 109:429-435, 1996; Chen et al., J. 5 Cell Biol. 138:891-899, 1997). It has been proposed that occludin promotes cell adhesion through homophilic interactions (an occludin on the surface of one cell binds to an identical occludin on the surface of another cell). A detailed discussion of occludin structure and function is provided by Lampugnani and Dejana, Curr. Opin Cell Biol. 9:674-682, 1997.

More recently, a second family of tight junction components has been identified. Claudins are transmembrane proteins that appear to be directly involved in cell adhesion and the formation of tight junctions (Furuse et al., J. Cell Biology 141:1539-1550, 1998; Morita et al., Proc. Natl. Acad. Sci. USA 96:511-516, 1999). Other previously described proteins that appear to be members of the claudin family include RVP-1 (Briehl and Miesfeld, Molecular Endocrinology 5:1381-1388, 1991; Katahira et al., J. Biological Chemistry 272:26652-26656, 1997), the Clostridium perfringens enterotoxin receptor (CPE-R; see Katahira et al., J. Cell Biology 136:1239-1247, 1997; Katahira et al., J. Biological Chemistry 272:26652-26656, 1997) and TMVCF (transmembrane protein deleted in Velo-cardio-facial syndrome; Sirotkin et al., Genomics 42:245-51, 1997).

Based on hydrophobicity analysis, all claudins appear to be approximately 22 kD and contain four hydrophobic domains that transverse the plasma membrane. It has been proposed that claudins promote cell adhesion through homophilic interactions (a claudin on the surface of one cell binds to an identical claudin on the surface of another cell) or heterophilic interactions, possibly with occludin.

Although cell adhesion is required for certain normal physiological functions, there are situations in which the level of cell adhesion is undesirable. For example, many pathologies (such as autoimmune diseases and inflammatory diseases) involve abnormal cellular adhesion. Cell adhesion may also play a role in graft rejection. In such circumstances, modulation of cell adhesion may be desirable.

In addition, permeability barriers arising from cell adhesion create difficulties for the delivery of drugs to specific tissues and tumors within the body. For example, skin patches are a convenient tool for administering drugs through the skin. However, the use of skin patches has been limited to small, hydrophobic molecules because of the epithelial and endothelial cell barriers. Similarly, endothelial cells render the blood capillaries largely impermeable to drugs, and the blood/brain

mammal, comprising administering to a mammal a cell adhesion modulating agent as provided above, wherein the modulating agent inhibits claudin-mediated cell adhesion.

The present invention further provides methods for inhibiting angiogenesis in a mammal, comprising administering to a mammal a cell adhesion modulating agent as provided above, wherein the modulating agent inhibits claudin mediated cell adhesion.

Within further aspects, the present invention provides methods for enhancing drug delivery to the central nervous system of a mammal, comprising administering to a mammal a cell adhesion modulating agent as provided above, wherein the modulating agent inhibits claudin-mediated cell adhesion.

The present invention further provides methods for enhancing wound healing in a mammal, comprising contacting a wound in a mammal with a cell adhesion modulating agent as provided above, wherein the modulating agent enhances claudin mediated cell adhesion.

Within a related aspect, the present invention provides methods for enhancing adhesion of foreign tissue implanted within a mammal, comprising contacting a site of implantation of foreign tissue in a mammal with a cell adhesion modulating agent as provided above, wherein the modulating agent enhances claudin mediated cell adhesion.

The present invention further provides methods for inducing apoptosis in a claudin-expressing cell, comprising contacting a claudin-expressing cell with a cell adhesion modulating agent as provided above, wherein the modulating agent inhibits claudin-mediated cell adhesion.

The present invention further provides methods for identifying an agent capable of modulating claudin-mediated cell adhesion. One such method comprises the steps of (a) culturing cells that express a claudin in the presence and absence of a candidate agent, under conditions and for a time sufficient to allow cell adhesion; and (b) visually evaluating the extent of cell adhesion among the cells.

Within another embodiment, such methods may comprise the steps of: (a) culturing normal rat kidney cells in the presence and absence of a candidate agent, under conditions and for a time sufficient to allow cell adhesion; and (b) comparing the level of cell surface claudin and E-cadherin for cells cultured in the presence of candidate agent to the level for cells cultured in the absence of candidate agent.

Within a further embodiment, such methods may comprise the steps of: (a) culturing human aortic endothelial cells in the presence and absence of a candidate agent, under conditions and for a time sufficient to allow cell adhesion; and (b) comparing the level of cell surface claudin and N-cadherin for cells cultured in the presence of candidate agent to the level for cells cultured in the absence of candidate agent.

Within yet another embodiment, such methods comprise the steps of: (a) contacting an

with other proteins of the junctional complex of the membrane skeleton (Gallagher and Forget, J. Biol. Chem., 270:26358-26363 (1995)). The proteins of SEQ ID NOs: 201 and 227 exhibit the PROSITE signature typical for the band 7 family signature.

The proteins of SEQ ID NOs: 201 and 227 play a role in the regulation of ion transport, hence in the control of cellular volume. These proteins are useful in diagnosing and/or treating stomatocytosis and/or cryohydrocytosis by detecting a decreased level or absence of the proteins or alternatively by detecting a mutation or deletion affecting tertiary structure of the proteins.

Protein of SEQ ID NO: 213 and 229

The proteins of SEQ ID NO: 213 and 229, encoded by the cDNA of SEQ ID NO: 166 and 228, respectively, is human Glia Maturation Factor-gamma 2 (GMF-gamma 2). SEQ ID NO: 229 differs from SEQ ID NO: 213 in that SEQ ID NO: 229 has additional amino acids at the N-terminus. The following description applies equally to both SEQ ID NO: 213 and 229. A preferred use of GMF-gamma 2 is to stimulate neurite outgrowth or neurite re-sprouting. These methods include both in vitro and in vivo uses, but preferred uses are those for treating neural injuries and cancer as disclosed in WO9739133 and WO9632959, incorporated herein in their entireties.

GMF-gamma 2 may also be used as a neurotrophic and as a neuroprotective agent against toxic insults, such as ethonal and other neurotoxic agents. GMF-gamma2 may be used as a neurotrophic or neuroprotective agent either in vitro or in vivo. A preferred target of GMF-gamma 2 as a neurotrophic or neuroprotective agent are primary neurons.

GMF-gamma 2 may further be used to stimulate the expression and secretion of NGF and BDNF in glial cells both in vitro and in vivo. Conditioned media from cells treated with GMF-gamma 2 is useful as a source of NGF and BDNF. GMF-gamma 2 may further be used to target cells directly or by recombinantly fusing GMF-gamma 2 to a heterologous protein, such as a ligand or antibody specific to the target cell (e.g., glial cells). Alternatively, GMF-gamma 2 may be fused or covalently or non-covalently coupled to a heterologous protein or other biological or non-biological molecule wherein the heterologous protein or molecule is used as this targeting reagent.

Preferred fragments of SEQ ID NOs: 213 and 229 and the corresponding polypeptide encoded by the human cDNAs of the deposited clones are those with the above activities. Further preferred are fragments with not less then 100 fold less activity, not less than 10 fold activity, and not less than 5 fold activity when compared to the protein of SEQ ID NO: 229 or the protein encoded by the corresponding human cDNA of the deposited clone.

Protein of SEQ ID NO: 214:

serine protease (Kato and Tominaga, Fed. Proc., 38:832 (1979)). All cysteines of the 4 disulfide core signature thought to be crucial for biological activity are present in the protein of SEQ ID NO: 216. The 4 disulfide core signature is present except for a conservative substitution of asparagine to glutamine.

Taken together, these data suggest that the protein of SEQ ID NO: 216 may play a role in protein-protein interaction, act as a protease inhibitor and/or may also be related to male fertility .

Protein of SEQ ID NO: 223

The protein of SEQ ID NO: 223 encoded by the extended cDNA SEQ ID NO: 176 shows homology to short stretches of a human protein called Tspan-1 (Genbank accession number AF054838) which belongs to the 4 transmembrane superfamily of molecular facilitators called tetraspanin (Meakers et al., FASEB J., 11:428-442 (1997)).

Taken together, these data suggest that the protein of SEQ ID NO: 223 may play a role in cell activation and proliferation, and/or adhesion and motility and/or differentiation and cancer.

As discussed above, the extended cDNAs of the present invention or portions thereof can be used for various purposes. The polynucleotides can be used to express recombinant protein for use for therapeutic use or research (not limited to research on the gene itself); as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support (e.g., microarrays), including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION

PATENT NO. : 6,573,068 B1
DATED : June 3, 2003
INVENTOR(S) : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert and Lydie Bougueleret



It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 3,
Line 62, "as "ESTs" or "5' EST"." should read -- as "ESTs" or "5'ESTs".--.

Column 16,
Line 8, "NOs. 2540 and 4246" should read -- NOs. 25-40 and 42-46 --.

Column 18,
Line 62, "5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCAACUCCUCCCAUC UCCAC-3'" should read --5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAA CUCCUCCCAUCUCCAC-3' --.

Column 22,
Line 61, "diameter 0.45 mn." should read -- diameter of 0.45 μm. --.

Column 24,
Line 67, "Caminci" should read -- Carninci --.

Column 28,
Line 8, "Insert" should read -- Inserts --.

Column 29,
Line 26, "helix-tum-helix" should read -- helix-turn-helix --.

Column 32,
Line 19, "14:46834690" should read -- 14:4683-4690 --.

Column 35,
Line 55, "270:467470" should read -- 270:467-470 --.

Column 41,
Line 43, "46834690" should read -- 4683-4690 --.

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 42,
Line 13, "proteins. 5. Selection of" should read -- proteins.
5. Selection of Cloned Full Length Sequencesof the Present Invention --.

Column 44,
Lines 2-3, "Table H" should read -- Table II --.
Line 7, "Table H" should read -- Table II --.

Column 57,
Line 50, "11:405411" should read -- 11:405-411 --.

Column 62,
Line 7, "13:473486" should read -- 13:473-486 --.

Column 63,
Line 59, "tendonaigament formation" should read -- tendon/ligament formation --.
Line 60, "tendonligament-like" should read -- tendon/ligament-like --.

Column 64,
Line 11, "tendonaigament cells" should read -- tendon/ligament cells --.

Column 65,
Line 9, "Chemokincs" should read -- Chemokines --.

Column 70,
Line 43, "Epitope" should read -- Epitopes --.

Column 74,
Lines 2-3, "5x $10^{-}{}_{15}$M, and" should read -- $5X10^{-15}$ M, and --.

Column 78,
Line 12, "*Bacic*" should read -- *Basic* --.

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION

PATENT NO. : 6,573,068 B1
DATED : June 3, 2003
INVENTOR(S) : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert and Lydie Bougueleret



It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 80,
Line *61*, "calorimetrically" should read -- colorimetrically --.

Column 85,
Line 9, "Genomics III:701-708" should read -- Genomics 11:701-708 --.

Column 86,
Line 63, "Extended cDNA" should read -- Extended cDNAs --.

Column 88,
Line 30, "Table VI" should read -- Table VII --.

Column 90,
Line 15, "3 min -94° C." should read -- 3 min -67° C. --.

Column 92,
Line 43, "Protein Which Internet" should read -- Proteins Which Interact --.

Column 98,
Line 25, "Table VI" should read --Table VII --.

Column 99,
Line 18, "412415" should read -- 412-415 --.
Line 19, "482488" should read -- 482-488 --.
Line 49, "Protein of" should read -- Proteins of --.

Column 100,
Line 35, "109:429435" should read -- 109:429-435 --.

Column 103,
Line 36, "E-cadhenn" should read -- E-cadherin -.

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION

PATENT NO. : 6,573,068 B1 
DATED : June 3, 2003
INVENTOR(S) : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert and Lydie Bougueleret

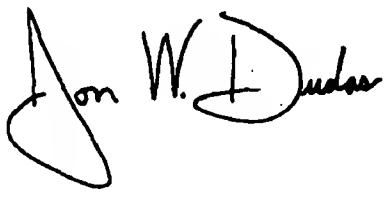
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 107,
Line 12, "both 15 in vitro and" should read -- both in vitro and --.

Column 108,
Line 46, "11:428442" should read -- 11:428-442 --.

Signed and Sealed this

Sixth Day of April, 2004



JON W. DUDAS
*Acting Director of the United States Patent and Trademark Office*

# UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



PATENT NO. : 6,573,068 B1

DATED : June 3, 2003

INVENTORS : Jean-Baptiste Dumas Milne Edwards, ~~Aymeric Duclert, Lydie Bougueleret~~

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 3,
Line 62, "as "ESTs" or "5' EST"." should read --as "ESTs" or "5'ESTs".--.

Column 16,
Line 8, "NOs. 2540 and 4246" should read --NOs. 25-40 and 42-46--.

Column 18,
Line 62, "5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCAACUCCUCCCAUC UCCAC-3'" should read --5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAA CUCCUCCCAUCUCCAC-3'--.

Column 22,
Line 61, "diameter 0.45 mn." should read --diameter of 0.45 μm.--

Column 24,
Line 67, "Caminci" should read --Carninci--.

Column 28,
Line 8, "Insert" should read --Inserts--.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606

PATENT NO. 6,573,068
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

OCT 1 6 2003

# UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



PATENT NO. : 6,573,068

DATED : June 3, 2003

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Lydie Bougueleret

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 29,
Line 26, "helix-tum-helix" should read --helix-turn-helix--.

Column 32,
Line 19, "14:46834690" should read --14:4683-4690--.

Column 35,
Line 55, "270:467470" should read --270:467-470--.

Column 41,
Line 43, "46834690" should read --4683-4690--.

Column 42,
Line 13, "proteins. 5. Selection of" should read --proteins.
5. Selection of Cloned Full Length Sequencesof the Present Invention--.

Column 44,
Lines 2-3, "Table H" should read --Table II--.
Line 7, "Table H" should read --Table II--.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606

PATENT NO. 6,573,068
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

# UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



PATENT NO. : 6,573,068

DATED : June 3, 2003

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Lydie Bougueleret

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 57,
Line 50, "11:405411" should read --11:405-411--.

Column 62,
Line 7, "13:473486" should read --13:473-486--.

Column 63,
Line 59, "tendonaigament formation" should read --tendon/ligament formation--.
Line 60, "tendonligament-like" should read --tendon/ligament-like--

Column 64,
Line 11, "tendonaigament cells" should read --tendon/ligament cells--.

Column 65,
Line 9, "Chemokincs" should read --Chemokines--.

Column 70,
Line 43, "Epitope" should read --Epitopes--.

Column 74,
Lines 2-3, "5x10-15M, and" should read --5X10$^{-15}$M, and--.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606

PATENT NO. 6,573,068
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

# UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



PATENT NO. : 6,573,068

DATED : June 3, 2003

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Lydie Bougueleret

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 78,
Line 12, "*Bacic*" should read --*Basic*--. C

Column 80,
Line 61, "calorimetrically" should read --colorimetrically--. C

Column 85,
Line 9, "Genomics II1:701-708" should read --Genomics 11:701-708-- C

Column 86,
Line 63, "Extended cDNA" should read --Extended cDNAs--. C

Column 88,
Line 30, "Table VI" should read --Table VII--. C

Column 90,
Line ~~43~~ 15, "3 min -94° C." should read --3 min -67° C.--. C

Note

Column 92,
Line 43, "Protein Which Internet" should read --Proteins Which Interact--. C

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606

PATENT NO. 6,573,068
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

# UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION



PATENT NO. : 6,573,068

DATED : June 3, 2003

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Lydie Bougueleret

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 98,
Line 25, "Table VI" should read --Table VII--.

Column 99,
Line 18, "412415" should read --412-415--.
Line 19, "482488" should read --482-488--.
Line 49, "Protein of" should read --Proteins of--.

Column 100,
Line 35, "109:429435" should read --109:429-435--.

Column 103,
Line 36, "E-cadhenn" should read --E-cadherin--.

Column 107,
Line 12, "both 15 in vitro and" should read --both in vitro and--.

Column 108,
Line 46, "11:428442" should read --11:428-442--.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
2421 N.W. 41st Street, Suite A-1
Gainesville, FL 32606

PATENT NO. 6,573,068
No. of add'l. copies
@ 30¢ per page